Project Report: Analyzing Hand Color for Health Status Prediction

Adli Karadsheh

August 02, 2025

Capstone 3

Introduction and Problem Statement

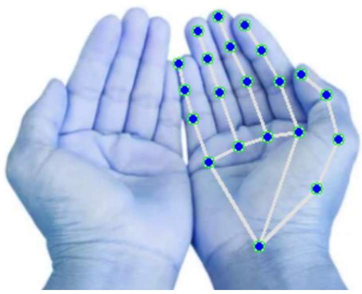
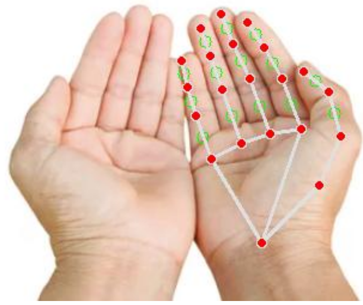
In healthcare, early detection of circulatory or oxygenation issues can significantly improve patient outcomes. One subtle indicator on physical exam of such problems is the color of a person's hands, where bluish hues may signal poor blood flow or low oxygen levels (hypoxia), while pink tones typically indicate healthy circulation. This project addresses the real-world problem of non-invasive health monitoring for at-risk individuals, such as those with cardiovascular conditions or respiratory disorders, by developing a machine learning model to classify hand images as "healthy" (pink tones) or "unhealthy" (bluish hues) based on color analysis.

The problem solution utilizes data science and machine learning approaches, leveraging data from hand images processed via MediaPipe for RGB feature extraction at midpoints between landmarks (to focus on interphalangeal areas). By automating this analysis, the model could be integrated into a mobile app for remote health screening, empowering users to seek timely medical advice.

Dataset and Data Sources

The dataset consists of hand images divided into two categories: "Healthy" (Label 1) and "Unhealthy" (Label 0). The "Healthy" folder contains ~70 images sourced from free stock photo websites (e.g., Pexels, Unsplash, Pixabay) and paid sources (Adobe Stock), in addition to photos obtained in the clinic, featuring hands with normal pink skin tones. The "Unhealthy" folder includes ~57 images obtained from an outpatient clinic, depicting hands with bluish hues indicative of potential health issues (e.g., chronic conditions affecting circulation).

Image Processing and Data Extraction

Images were processed using MediaPipe Hands to detect landmarks and extract RGB values at midpoints of finger segments (thumb: 2 segments; others: 3), yielding ~84 features per image (RGB × segments × up to 2 hands). Sampling radii of 1, 3, and 5 pixels were applied to create three CSVs (`hand_color_data_midpoints_radius_{radius}.csv`), with NaNs (from boundary issues or single-hand images) imputed via column means. No rows were lost to imputation. The dataset is mildly imbalanced (~58% unhealthy), addressed via stratified splits in modeling. T-tests confirmed significant RGB differences ($p < 0.001$ across channels/radii, green strongest t ~5.9-6.0).

Radius	Channel	t-statistic	p-value
1	Avg_R	4.22	4.7240e-05
1	Avg_G	5.90	3.1813e-08
1	Avg_B	4.99	1.9682e-06
3	Avg_R	4.15	6.1038e-05
3	Avg_G	5.89	3.3506e-08
3	Avg_B	4.96	2.2910e-06
5	Avg_R	4.11	7.1264e-05
5	Avg_G	5.99	2.0721e-08
5	Avg_B	5.06	1.4491e-06

Data Wrangling and Exploratory Data Analysis (EDA)

- Wrangling: Loaded CSVs, imputed NaNs (means), added average RGB columns (Avg_R/G/B). No dummy features needed (numeric RGB). Standardized magnitudes via StandardScaler. Split 80/20 train-test (stratified, random_state=42).

- EDA (`EDA_radii135_datasets.ipynb`):

  - Shape: 127 rows, ~86 columns per radius.

  - Distributions: RGB ~50-255 (skin tones); healthy brighter (means ~217R/169G/149B vs. unhealthy ~206R/142G/132B at R1).

  - Correlations: High within channels (~0.9+), moderate with Label (~0.4-0.5, red/green strongest).

  - Visuals: Avg RGB bars (healthy taller, green widest gap); histograms (bimodal for labels); boxplots (healthy higher medians, fewer outliers); pairplots (separation in R-G); class balance (~58/42 unhealthy). Larger radii smooth data (lower std, better separation). Outliers minimal (~0-2 low RGB in unhealthy).

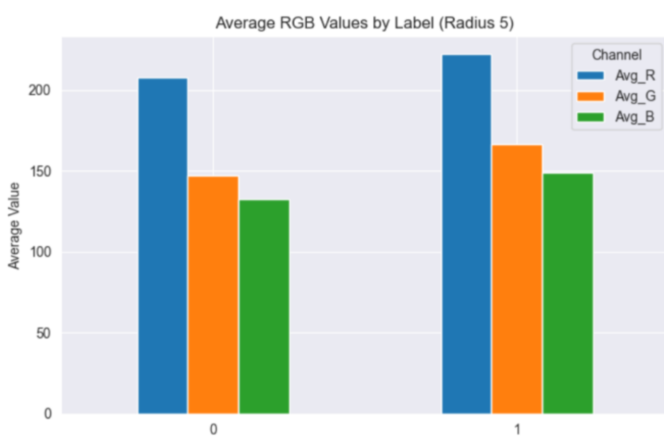


Figure 1: Healthy hands show higher RGB values, especially green.

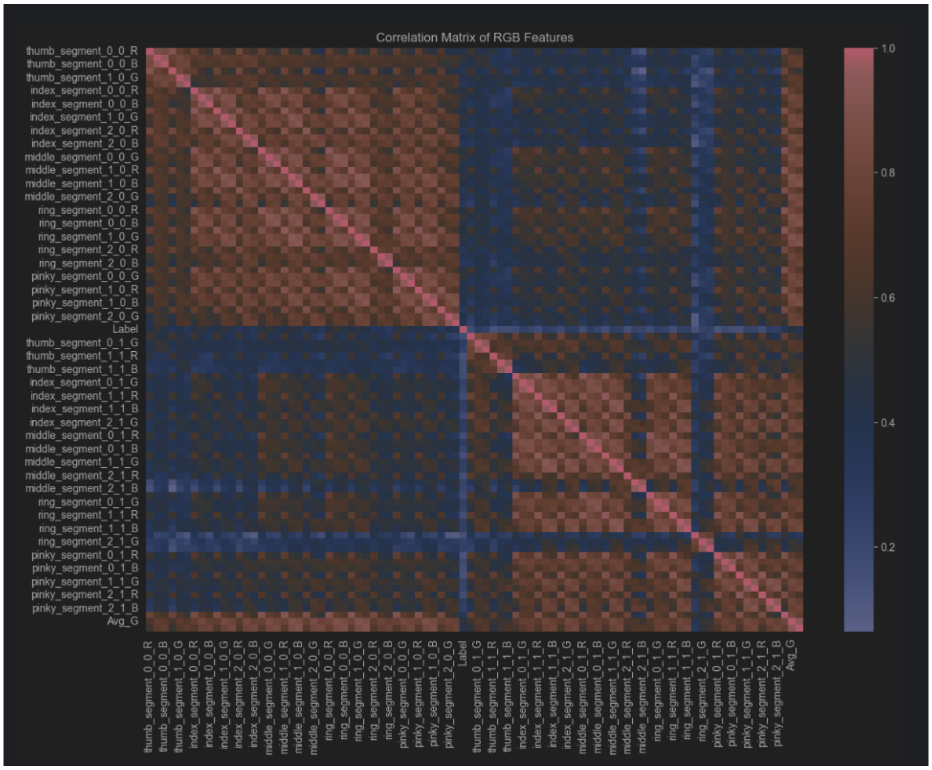


Figure 2: Strong within-channel correlations; Label positive with RGB.

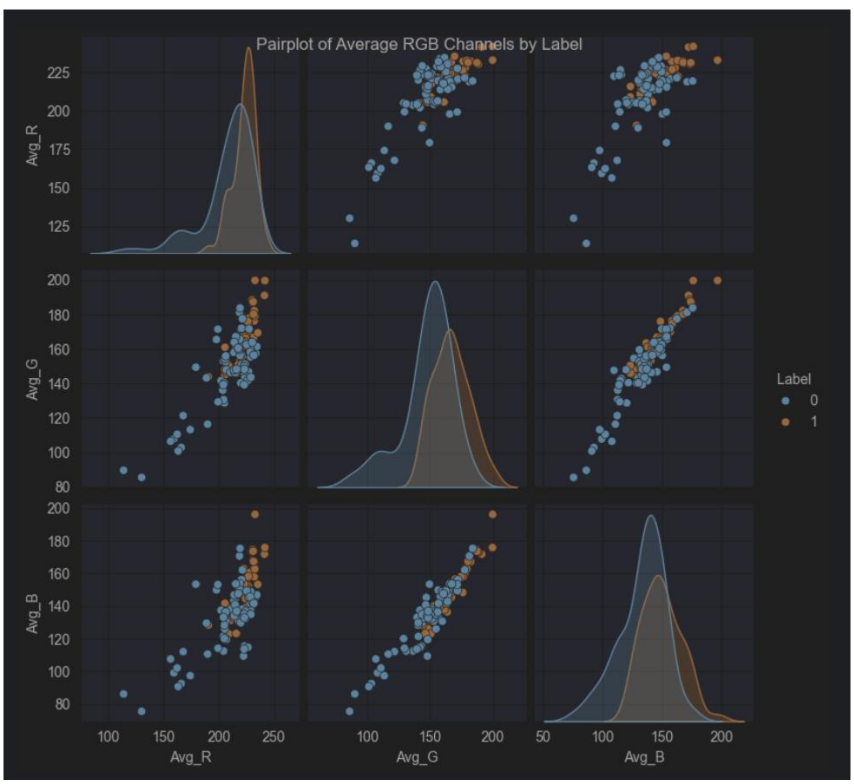


Figure 3: Clear clusters; minimal overlap in R-G.

Modeling

Four models were trained per radius (`Hands_ML{1-4}.ipynb`), using PCA (20 components, ~97% variance) for reduction:

- Initial Logistic Regression (LR): Baseline; test ~81-88%, CV ~68-71% (±0.11-0.12). Warnings from separation/collinearity.
- Regularized LR (L2, C=1.0): Stabilized; test ~85-88%, CV ~72-74% (±0.10-0.12). Coeffs highlight Comp 16 (green contrasts).
- Random Forest (RF, n_estimators=100): Best; test 92% across, CV 72-83% (±0.10-0.12). Importances: Comp 1 (~12%, global shifts), Comp 16 (~9% at R5).
- Neural Network (NN, 64-32-1 layers, 50 epochs): Test ~81-88%, CV ~74-76% (±0.06-0.12, averaged over runs for stochasticity).

Final model: RF at R5 (92% test, 83% CV)—ensemble captures non-linearities; low FN (1) prioritizes unhealthy detection.

Model Comparison (`model_comparison.csv` averages):

Model	Radius	Test Acc	CV Acc (± std)	CM Insights (R5)
Initial LR	5	0.81	0.71 (±0.11)	TN=12/15, 3 FP, 2 FN
Regularized LR	5	0.85	0.72 (±0.10)	TN=13/15, 2 FP, 2 FN
Random Forest	**5**	**0.92**	**0.83 (±0.12)**	**TN=14/15, 1 FP, 1 FN**
Neural Network	5	0.83	0.76 (±0.06)	TN=13/15 avg, 2 FP/FN

Assessment of Predictions vs. Actual Outcomes

Predictions were evaluated on test sets (26 samples per radius). RF is superior with near-perfect alignment (92% correct; e.g., R5 CM: 14/15 unhealthy correct, only 1 FN—misses few cases). Errors: Rare FP/FN often in borderline images (e.g., lighting artifacts per EDA outliers). Vs. actual: Healthy (pink) well-predicted (high TP); unhealthy (blue) prioritized (low FN). NN/Reg LR show more misalignment (2 FN, higher FP at R5 due to smoothing). Radius trend: R5 best (smoothing reduces false positives from noise).

RF at R5 provides reliable health prediction (92% accuracy), confirming RGB midpoints as effective features. Future: App integration, larger datasets.

Conclusion

This capstone project demonstrates the viability of non-invasive hand color analysis for binary health classification, with Random Forest at radius 5 pixels emerging as the top performer (92% test accuracy, 83% CV). By focusing on midpoint RGB sampling, the approach effectively captures circulatory indicators—healthy hands exhibit brighter tones (e.g., Avg_G ~160 vs. ~140 for unhealthy, $p < 0.001$)—paving the way for accessible remote screening tools. However, limitations include the small dataset (~127 samples per radius), risking overfitting (evident in NN's CV variability ±0.06-0.12) and generalizability; multicollinearity in RGB features (~0.9+ correlations) may inflate variance; and unaccounted factors like diverse skin tones or lighting could bias results (e.g., outliers in unhealthy low RGB likely from shadows). Future validation on larger, diverse cohorts is essential.

Next steps to advance this work:

1. **Enhance Photo Processing**: Incorporate HSV color space (Hue for tone consistency, Saturation/Value for brightness/illumination) alongside RGB to improve robustness—e.g., HSV's Value channel could normalize lighting variations. Standardize images to uniform sizes (e.g., 512x512 via cv2.resize) and apply corrections (e.g., histogram equalization or white balance using OpenCV) to mitigate environmental factors like ambient light, tested via A/B comparisons on accuracy.
2. **Refine Labeling**: Augment metadata with clinical details (e.g., age, chronic conditions like diabetes/COPD, oxygen saturation via pulse oximetry) to enable multi-class or

regression models. This could involve collaborating with clinics for annotated data, improving interpretability (e.g., correlate bluish hues with specific conditions) and addressing ethical biases (e.g., age/skin tone fairness).

3. **Optimize and Compare Models**: Tune hyperparameters (e.g., RF n_estimators=200, NN dropout=0.2 for variance reduction) using GridSearchCV; explore ensembles (RF+NN) or advanced architectures (CNNs like ResNet on raw images). Compare via expanded metrics (precision/recall for FN minimization, SHAP for feature explanations)—aiming for >95% CV on augmented data.

4. **Develop an App**: Prototype a mobile app (Flutter/React Native) integrating the RF model (via TensorFlow Lite/ONNX export) for real-time camera input, with user-friendly outputs (e.g., health score + alerts). Include privacy features (on-device processing) and pilot testing for usability/clinical validation.

These steps could evolve the model into a practical tool, bridging data science with healthcare impact.

References: MediaPipe Docs, Scikit-learn/TensorFlow Tutorials, Pexels/Unsplash.